Original Paper

# Issues in Building a Nursing Home Syndromic Surveillance System with Textmining: Longitudinal Observational Study

Tiba Delespierre[*], MSc, PhD; Loic Josseran[*], MD, PhD

Equipe de recherche (HANDIReSP), UFR des Sciences de la Santé Simone Veil, Université de Versailles Saint-Quentin-en-Yvelines et Université Paris-Saclay, Montigny-le-Bretonneux, France

[*]all authors contributed equally

**Corresponding Author:**
Tiba Delespierre, MSc, PhD
Equipe de recherche (HANDIReSP)
UFR des Sciences de la Santé Simone Veil
Université de Versailles Saint-Quentin-en-Yvelines et Université Paris-Saclay
2 Avenue de la Source de la Bièvre
Montigny-le-Bretonneux, 78180
France
Phone: 33 658376503
Fax: 33 170429515
Email: tiba.baroukh@gmail.com

## Abstract

**Background:** New nursing homes (NH) data warehouses fed from residents' medical records allow monitoring the health of elderly population on a daily basis. Elsewhere, syndromic surveillance has already shown that professional data can be used for public health (PH) surveillance but not during a long-term follow-up of the same cohort.

**Objective:** This study aimed to build and assess a national ecological NH PH surveillance system (SS).

**Methods:** Using a national network of 126 NH, we built a residents' cohort, extracted medical and personal data from their electronic health records, and transmitted them through the internet to a national server almost in real time. After recording sociodemographic, autonomic and syndromic information, a set of 26 syndromes was defined using pattern matching with the standard query language-LIKE operator and a Delphi-like technique, between November 2010 and June 2016. We used early aberration reporting system (EARS) and Bayes surveillance algorithms of the R surveillance package (Höhle) to assess our influenza and acute gastroenteritis (AGE) syndromic data against the Sentinelles network data, French epidemics gold standard, following Centers for Disease Control and Prevention surveillance system assessment guidelines.

**Results:** By extracting all sociodemographic residents' data, a cohort of 41,061 senior citizens was built. EARS_C3 algorithm on NH influenza and AGE syndromic data gave sensitivities of 0.482 and 0.539 and specificities of 0.844 and 0.952, respectively, over a 6-year period, forecasting the last influenza outbreak by catching early flu signals. In addition, assessment of influenza and AGE syndromic data quality showed precisions of 0.98 and 0.96 during last season epidemic weeks' peaks (weeks 03-2017 and 01-2017) and precisions of 0.95 and 0.92 during last summer epidemic weeks' low (week 33-2016).

**Conclusions:** This study confirmed that using syndromic information gives a good opportunity to develop a genuine French national PH SS dedicated to senior citizens. Access to senior citizens' free-text validated health data on influenza and AGE responds to a PH issue for the surveillance of this fragile population. This database will also make possible new ecological research on other subjects that will improve prevention, care, and rapid response when facing health threats.

XSL•FO
**RenderX**

# Introduction

## Background

Population in developed countries is aging [1], and the French population follows this trend. In France, by 2050, 22.3 million people will be aged over 65 years compared with 12.6 million in 2005, an increase of 80% in 45 years. Between 2013 and 2050, the senior population will grow more than the population as a whole. Similarly, life expectancy at birth in France, one of the highest in the world, is projected to surpass 86 years for men and 90 for women [2].

This increase will then have to be anticipated and will affect care and related costs [3]. It is, therefore, essential to improve our knowledge of this senescence process to help prevent increase in pathologies, and improve quality of life at extreme ages.

In spite of this major expected evolution of population, ecological research on this aged population is still limited [4]. Case or ad hoc studies do not consider individual variability and cannot analyze health issues as a whole. Data then need to be recorded for quite a long time and on a daily basis, helping to address this lack of knowledge. This has to be done in a natural way, in a professional environment with caregivers and medical staff [5].

As until now, follow-up studies on senior citizens were conducted using cohorts that were costly to set up and follow [6-8]. Data are occasionally stored, even if the follow-up is long and based on auto-questionnaires or planned interviews with health professionals. This approach does not allow describing in detail the daily life of this population and storing health evolutions throughout the residents' whole stays.

## New Data

On the contrary, nursing homes (NH) offer this possibility of tracking and recording them daily as health professionals feed these information for their proper use and, this time, without any memory bias [9]. These new data as well as their uses suggest innovative approaches to improve health knowledge.

Korian (Paris, France) as the first private NH European group has these kinds of data. This enterprise holds 290 NH and approximately 3.92% (290/7394) [10] of the French NH network, distributed all over the country, mostly in urban areas (see Multimedia Appendix 1). A professional data warehouse (DWH) set up in 2010 hosts half the company's French residents' population data. Their health follow-ups are recorded daily from 126 NH. For every new resident admitted in one of the NH, a personal electronic resident medical file (PERMF) is set up. Data are then collected at various times: at admission (admission date, medical history, marital status, birth date, tastes, and habits), on a daily basis (new pathologies, chronic disease evolution, date of death, and drug prescriptions), or just after specific medical or health care professional visits. Items include diagnosis, outcomes, as well as sociodemographic information.

## Objectives

Elsewhere and a little earlier, during the early 2000s, surveillance systems (SS) [11-19] showed that professional data

could also be used for health and alert surveillance [20-26]. Here, professional data use for SS was only done using point data analysis (going to the emergency, 911, and Web queries) [19,25,26] and not during a long-term follow-up of the same people, and even more, not dedicated to senior citizens.

As we have just seen, data gathered by different NH professionals offer the opportunity of following the residents' situations on the flow and on a daily basis and, through this process, of building syndromic surveillance data. The objective of this study was to build and assess an influenza and acute gastroenteritis national ecological NH public health (PH) SS describing and validating the Base du Bien Vieillir (BBV), that is, Aging Well database architecture. Thus, through a new health data building paradigm, we engineered an NH syndromic surveillance system (SSS) based on already validated criteria [11,13], hopefully opening the way to new research and knowledge about the senescence process.

# Methods

## Data Collection

All data are transmitted from 126 NH in real time to a national server using the group intranet. Records collected from the PERMF server are anonymized (see Multimedia Appendix 2) when sent to the BBV server, keeping track of every resident even when moving from one NH to another. After this first step, BBV is built through an extract, transform, and load (ETL) process of health and sociodemographic data. (see Multimedia Appendix 3 for details). Following this second step, all residents have two types of data:

1. Gender, age, and GIR (Groupe ISO Ressources; english: Group International Standardization Organization Resource Group), a French autonomy-level rating indexed to government benefits [27-32] at the NH entry
2. Daily care information fed on the flow by the caregivers and the medical staff, whenever deemed useful, that is, their syndromic information and, finally, hospitalizations and death

At the same time, every Sunday, all residents' daily care information is aggregated to count the weekly number of syndromes per NH.

By extracting data for all residents of the PERMRF database from its inception, from November 1, 2010, to mid-February 2017, and adding every new resident entering one of the NH networks every week, a *one-week moving* cohort of residents followed during their entire NH life course was built opening the way for our SSS. Even if most residents of this cohort were followed during their entire NH life course, syndromic data could be left-truncated for people whose data were entered before the inception of the information system (IS) or right-truncated for people whose data were entered later.

At the IS core, the data transmissions table containing key information about the residents' care was fed on a daily basis. Data take the form of big size character fields (of up to 4000 characters). Extracting these and using residents and NHs' indexes and data transmissions dates (see Multimedia Appendix

XSL•FO

**RenderX**

3 for a complete example), we were able to track all residents through two dimensions. Over time—every day with syndromic data—and space—every NH with syndromic data—with queries and text mining, building their syndromic life course, beginning at their date of entry and ending with their last available data transmission or death. The BBV then has two nested time frames: by day for every resident and by week for every NH.

## Ethics Approval and Consent to Participate

The use of this database in the frame of epidemiological studies has been authorized by the French National Commission for Data protection and Liberties. The Institut du Bien Vieillir, which became the Foundation Korian of Well Ageing, filed a declaration of conformity to a baseline methodology, which received an agreement number in March 2017: 2.041.050, in accordance with the Act n 78–17 of January 6, 1978 on Data Processing, Data Files, and Individual Liberties. All residents are informed at their NH entry about their electronic health record (EHR) and their right to oppose its use. Although the primary purpose of this medical research was to generate new knowledge, this goal did not take precedence over the rights and interests of the NH residents. All the new generated information was extracted from already existing data and was deidentified and anonymized when necessary to protect their health and rights.

## Building the Acute Respiratory Infection and Influenza-Like Illness and Acute Gastroenteritis Syndromes

With a multistep learning and text mining (MSL-TM) process (see Multimedia Appendix 3 for the 4-phase process) of the data transmissions file similar to what was experimented in the study by Cohen at al [33], using problems' list logic [34-36] and pattern matching with the SQL-like operator [37], 24 syndromes were implemented [38-49], following the SurSaUD (Sanitary Surveillance of Urgencies and Deaths) SSS method [16]. Starting with acute respiratory infection and influenza-like illness (ARI-ILI) and acute gastroenteritis (AGE) syndromes (see Multimedia Appendix 3 for 2 examples and the syndromes' list), extracting directly hospitalizations and deaths, this NH IS kept for every resident, every day, in every NH, from none to 26 daily syndromes whenever appropriate (see Multimedia Appendix 3 again for full details of the whole process [50-61]).

## The Surveillance Tools Framework

Syndromic Systems attempt to detect outbreaks through statistical analysis of aggregated cases data to improve on competent clinicians in detecting early-stage or small outbreaks [62]. It focuses on data collected before clinical diagnosis or laboratory confirmation [63]. Statistical laws are then defined to give an answer to the question "knowing the average number of expected events during a period of time, what is the probability to observe the current situation?" [62].

The SSS generation was designed using a *Pentaho* extraction platform for all the ETL processes [64] and is described in Figure 1. It follows the Centers for Disease Control and Prevention (CDC) Working Group recommendations [11,13].

The whole process was done in 4 steps: first, the ILI and AGE syndromes built through the MSL-TM process [65]; second, the weekly ILI and AGE syndromic data aggregation and the time series (TS) generation with their statistical alerts using the R *surveillance* package [66,67]; third, the Sentinelles data joining, the ARI-ILI and AGE French surveillance gold standard [68]; and finally, the alerting system interfacing the *surveillance* package [66,67] statistical alerts with the NH general practitioners (GPs) coordinators signals, eventually reporting to the Health Regional Agencies (HRA).

It is only after the final step that epidemiologists in the national public health agency's regional units (HRA in Figure 1) are asked to choose an alert level for the regions they are in charge of: non-epidemic, pre/post epidemic, or epidemic [68]. A public health alert will then be defined as such by the public health agency Santé publique France (SPF) after every signal has been verified and validated [69] (for further details, see Multimedia Appendix 4 [22,67-76]). Relevant information for French epidemiologists since January 2016 at a regional level includes the Sentinelles (2.1% of French private GP) as well as the OSCOUR (Organization of the COordinated Surveillance of Urgencies; 88% of French hospital emergency departments make up the Coordinated Health Surveillance of Emergency Department network) and SurSaUD (95% of French emergency GP consultations) data but also local specific surveillance data such as NH ARI clusters' surveillance.

## Syndromic Data Analysis

### *Data Flow Buildup and Stabilization*

As explained above, we computed weekly counts of ILI and AGE cases as well as hospitalizations and deaths as with *Sentinelles*, avoiding week and weekend days' heterogeneity [77]. Then, with the *ggplot* function of the *ggplot2* R package [78] used with local regression curves fitted to the NH data (Figure 2) [79], we were able to track yearly tendencies as well as inconsistent data not reflecting the seasonal spikes during winter.

Assessing the syndromic data flow over time, by computing the summary statistics of deaths, hospitalizations, ARI-ILI and AGE weekly syndromes' counts during the 3 periods (ie, from November 1, 2010, to November 1, 2011; then from November 1, 2011, to November 1, 2012; and finally from November 1, 2010, to February 26, 2017), we chose to exclude the first year's data from this analysis (see Table 1).

**Figure 1.** The nursing homes acute respiratory infections and influenza-like illness and acute gastroenteritis surveillance tools framework. BBV: Base du Bien Vieillir (ageing well database); ILI: influenza-like Illness; AGE: acute gastroenteritis; MSL: multi-step learning; TM: text mining; TS: time series; GP: general practioner; HRA: Health Regional Agencies.
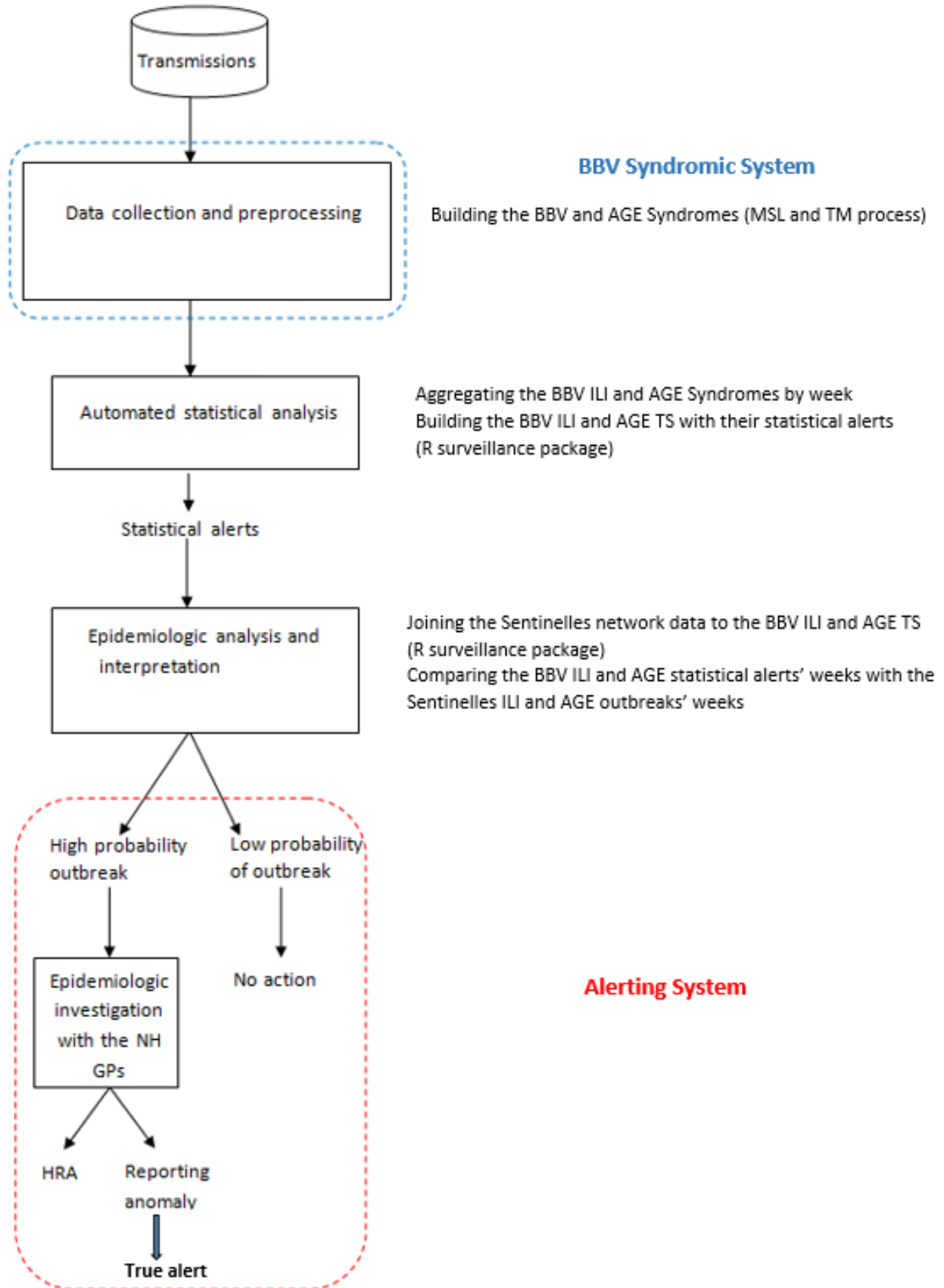
**Figure 2.** Acute gastroenteritis, acute respiratory infections, deaths and hospitalizations data flow buildup and stabilization in 11 regions covering France between November 1, 2010, and February 26, 2017. AGE-DIAR: acute gastroenteritis and diarrhea; ARI-ILI: acute respiratory infection and influenza-like illness.
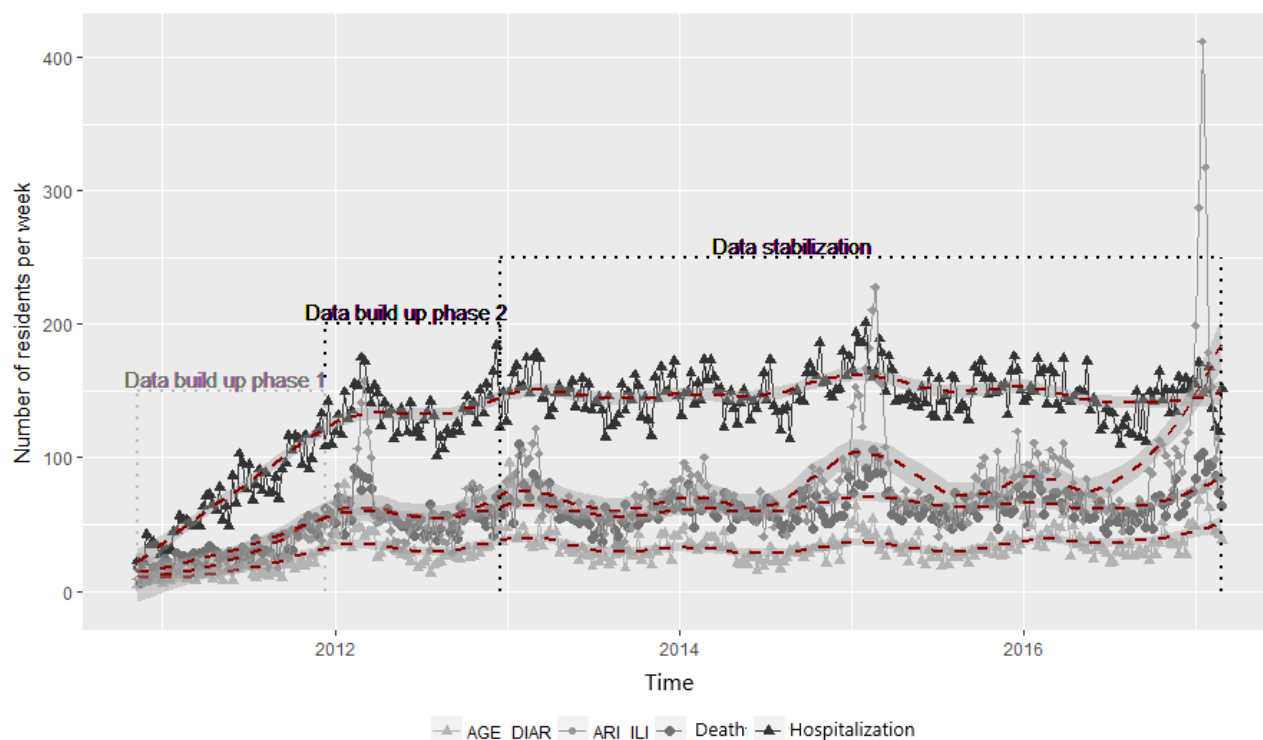


**Table 1.** Assessing the Base du Bien Vieillir (BBV) four syndrome's weekly counts.

| Period | Minimum | Q1 | Median | Mean | Q3 | Maximum |
|---|---|---|---|---|---|---|
| Phase 1 | 4 | 16 | 24.5 | 33.45 | 42.25 | 118 |
| Phase 2 | 13 | 40.5 | 55 | 70.92 | 109.25 | 184 |
| Stabilization | 15 | 48 | 67 | 82.72 | 124 | 412 |

### Building the Acute Respiratory Infection and Influenza-Like Illness and Acute Gastroenteritis Time Series

The ARI-ILI and AGE TS were built by aggregating all NH weekly syndromes counts. The choice of a statistical method to analyze these then rested on definitions of statistical alerts, adapting the BBV SSS data to fit French public health infrastructure and available SS data sources [80], here data of the ARI-ILI and AGE Sentinelles network [22]. Although the Sentinelles network used the Serfling method [70,71] relying on disease incidence levels of preceding years, we used the CDC steady favorite, the CUSUM (cumulative sums) methods, not drawing on data from preceding years but just from preceding weeks and 1 recent method, the Bayes method allowing fine tuning [72]. For further details, see Multimedia Appendix 4.

### Quality-Precision

Following the process described in the Building the Acute Respiratory Infection and Influenza-Like Illness and Acute Gastroenteritis Time Series subsection, the whole procedure was reviewed over 3 weeks of data transmissions: one in

mid-August 2016 when there was no epidemic and two in January 2017 at the ILI and AGE epidemics weeks' peaks, respectively, according to the Sentinelles network [81], computing the percentage of miscoded ARI-ILI and AGE syndromes among the extracted data transmissions defined as such [26].

### Stability

The idea here was to check the syndromic flow stability in quantity (the syndromes counts) and quality (several different recurring syndromes) during the complete period and for all 126 NH, computing the weekly syndromes frequencies for every NH.

The syndromic data flow stability was traced by designing 3 chronic diseases and 1 often-chronic ailment indexes [82] built as follows: whenever a resident had diabetes or a cardiovascular problem or depression or fell, the resident's transmission date and syndromic event type were set apart. Then, a similar event during a 200-day period after this resident's syndromic event was searched for, defining 4 syndromic ratios for the 6 years from the year 2011 up to February 27, 2017. For further details, see Multimedia Appendix 5.

### *Flexibility-Timeliness-Representativeness-Usefulness*

Adaptability and reactiveness of the system were evaluated during the outbreak and routine periods according to the CDC surveillance systems guidelines [18,15,26]. Representativeness, completeness, and usefulness were assessed using the distribution description of ILI cases by time and origin during this last flu season, as well as by rating sex, age, and GIR at entry and age at illness missing data [13].

### *Surveillance Algorithms' Quality*

All surveillance methods involve building first, time series with the weekly number of cases and second, statistical indicators used as thresholds. Here the 4 algorithms were compared by using the *algo.quality* function for Bayes [72] and rebuilding it for the early aberration reporting system (EARS) algorithms. This quality is defined by 4 numbers—the number of true positive (TP), false positive (FP), true negative (TN), false negative (FN)—and 4 criteria—the *sensitivity* [83] sometimes called recall [84] as the ratio of epidemic weeks correctly identified; the *specificity*, as the ratio of nonepidemic weeks correctly identified; the Euclidean distance between the perfect method with *specificity=sensitivity=1* and ours ($distance=((1-specificity)^2+(sensitivity-1)^2))^{1/2}$; and finally, the *precision* or *positive predictive value* (PPV), as the ratio of epidemic weeks correctly identified among the weeks defined as epidemic (with statistical alarm) [13].

# *Results*

## The Cohort

As explained in the Data Collection subsection, by extracting all residents already there on November 1, 2010, and then by adding those entered every week in one of the 126 NH, a cohort of 41,061 residents (Figure 2) was built with 12,983 men (31.61%, 12,983/41,061) of mean age 84.33 years and 28,083 women (68.39 % 28,083/41,061) of mean age 85.82 years.

### *The Acute Respiratory Infection and Influenza-Like Illness and Acute Gastroenteritis Syndromes and the Surveillance Inside Korian*

As described in Building the Acute Respiratory Infection and Influenza-Like Illness and Acute Gastroenteritis Syndromes subsection, the BBV syndromic algorithm extracted all ARI-ILI and AGE cases in addition to hospitalizations and deaths, every week, from November 1, 2010, to mid-February 2017 and built the 4 TS. Using the BBV ARI-ILI and AGE syndromic TS, we were able to track the last flu season (winter 2016-2017) early on, even before the epidemic and compare our syndromic counts with the Korian GPs' number of cases. The first ones were usually much greater than the second ones, as several syndromic cases could identify the same resident over time, but both of them were always strongly correlated.

## Syndromic Data Analysis

### *Checking the Data Flow During Time*

We managed to highlight 3 different phases in the NH data flow as shown in Figure 2 and Table 1 with two buildup phases during the first 2 years of the IS implementation. As seen below, between the first and the second year, the median and mean weekly syndromes' counts more than doubled. For that reason, we excluded the data of the first year from the syndromic data analysis.

### *The Acute Respiratory Infection and Influenza-Like Illness and Acute Gastroenteritis Time Series*

In the *surveillance* package, both Bayes' (see Figures 3 and 4) and EARS_C3s' algorithms with alpha=.025 (see Figures 5 and 6) used the 12 former ARI_ILI (Figures 3 and 5) and AGE (Figures 4 and 6) NH weeks' counts to define *alarm weeks* (red triangles). *Outbreak weeks* (green vertical lines) were defined according to the ILI and AGE Sentinelles data during the same period (from January 1, 2011, to January 16, 2011). Finally, the blue dotted lines were the upper limits at which alarms were triggered with both algorithms.

Senior citizens suffer much more of either ARI-ILI or AGE than the general population all year long and even in summer. This often results mechanically in triggering statistical alerts long before the general population epidemics. It can be seen in Figures 3 and 5 for ARI_ILI and Figures 4 and 6 for AGE where the red triangles (the SSS alarm weeks) appear always before the green bars (the Sentinelles' network outbreak weeks). It is especially true for ARI-ILI during the 2013 to 2014 and 2015 to 2016 winters and for AGE during the 2013 to 2014 winter with both algorithms.

### *Quality-Precision*

To compute the percentage of miscoded ARI-*ILI* and *AGE* syndromes, all ARI-ILI and AGE syndromic data were extracted during 3 weeks: one in mid-August 2016, the 33rd week (third column) when there was neither flu nor AGE epidemic, and two in January 2017, the third and first weeks, at the ILI and AGE epidemics weeks' peaks, according to the Sentinelles network (second column). Then, each ILI and AGE syndromic data transmission was examined, rating it as correct, adding to TP, or incorrect, adding to FP (see Table 2).

The precision was best during epidemic weeks' peaks: 98% and 96%, respectively, as ILI and AGE versus 95% and 92% in summer, and there were very little FP. For example, for ILI FP, " his son has flu," "emergencies overloaded with flu cases," "no flu symptoms," "could they take care of my girl who has flu?", and finally, "serrure dégrippée," which means *unjammed* in French but has the same word stem "gripp" as flu, were excluded. We had already excluded the word " grippé" in this context, which means *jammed* for a lock. In addition, by checking the flu cases, we found 67 flu tests mentions using nasal swabs, adding 21% new cases. For AGE FP, "vomited without diarrhea" (2 times, as both words are needed to classify as an AGE syndrome) and "diarrhea protocol if fever" were excluded.

**Figure 3.** The influenza-like illness (ILI) Base du Bien Vieillir time series (TS) using the Bayes' alarm algorithm with 12 weeks upstream and the ILI Sentinelles outbreaks. NH: nursing homes. Green ellipses highlight a nice overlapping of alarm and Sentinelles' network weeks or when the algorithm seems better, whereas red ellipses when this is not the case.
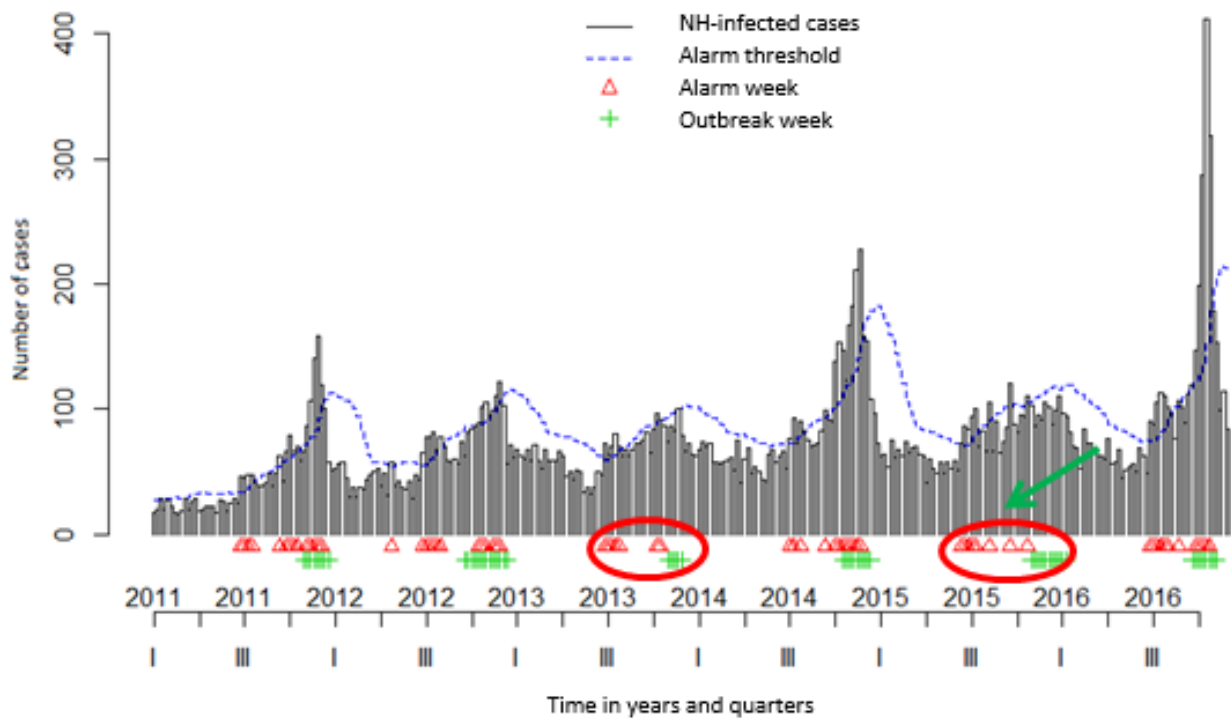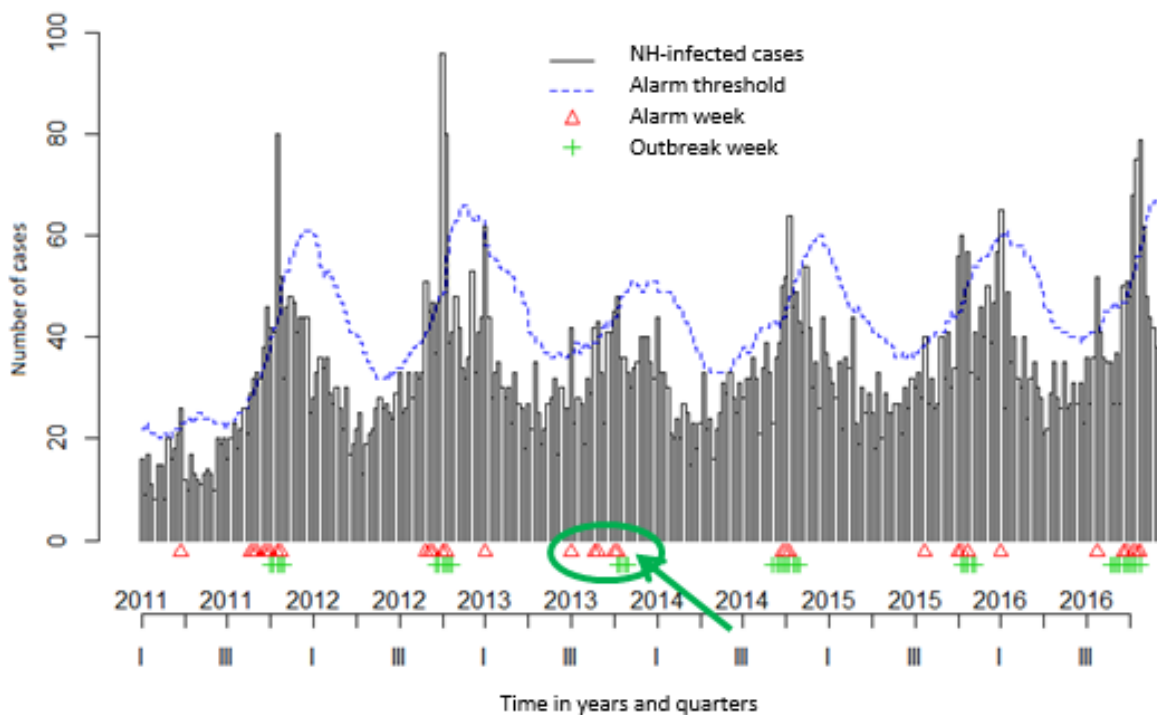


**Figure 4.** The acute gastroenteritis (AGE) Base du Bien Vieillir (BBV) time series using the Bayes' alarm algorithm with 12 weeks upstream and the AGE Sentinelles outbreaks. NH: nursing homes. Green ellipses highlight a nice overlapping of alarm and Sentinelles' network weeks or when the algorithm seems better, whereas red ellipses when this is not the case.
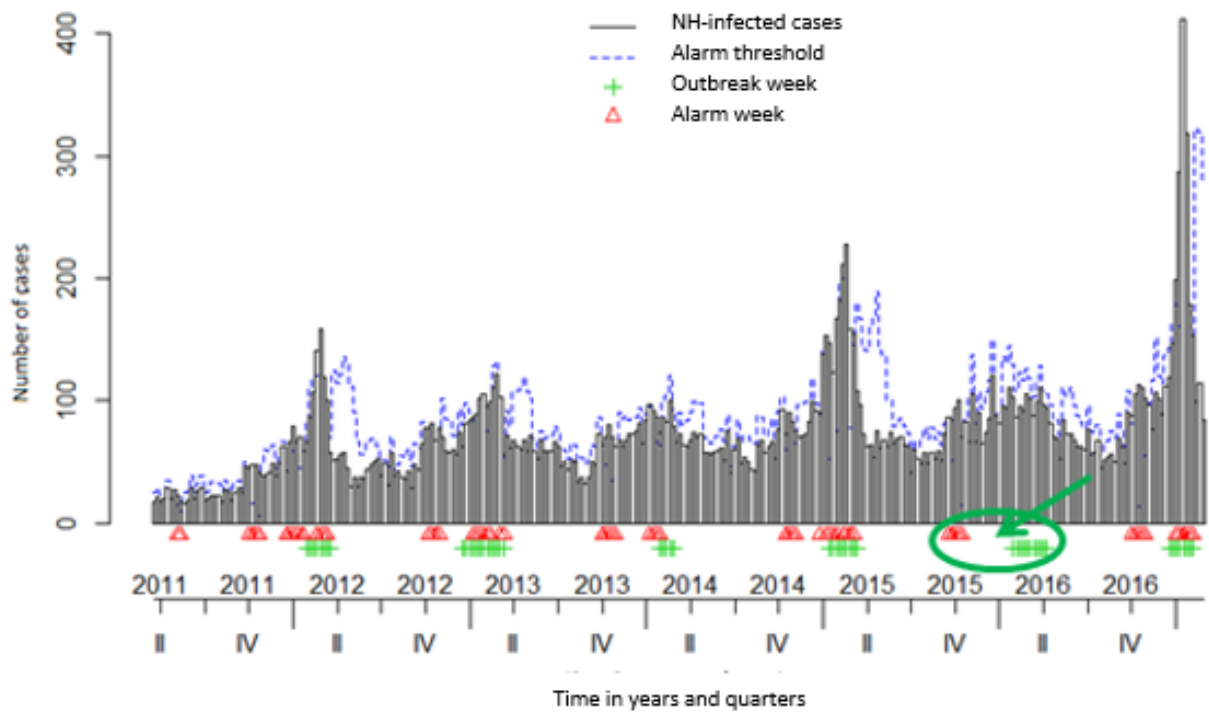
**Figure 5.** The influenza-like illness (ILI) Base du Bien Vieillir (BBV) time series using the EARS_C3's alarm algorithm with 12 weeks upstream and the ILI Sentinelles outbreaks. NH: nursing homes. Green ellipses highlight a nice overlapping of alarm and Sentinelles' network weeks or when the algorithm seems better, whereas red ellipses when this is not the case.

**Figure 6.** The acute gastroenteritis (AGE) Base du Bien Vieillir (BBV) time series using the EARS_C3's alarm algorithm with 12 weeks upstream and the AGE Sentinelles outbreaks. NH: nursing homes. Green ellipses highlight a nice overlapping of alarm and Sentinelles' network weeks or when the algorithm seems better, whereas red ellipses when this is not the case.
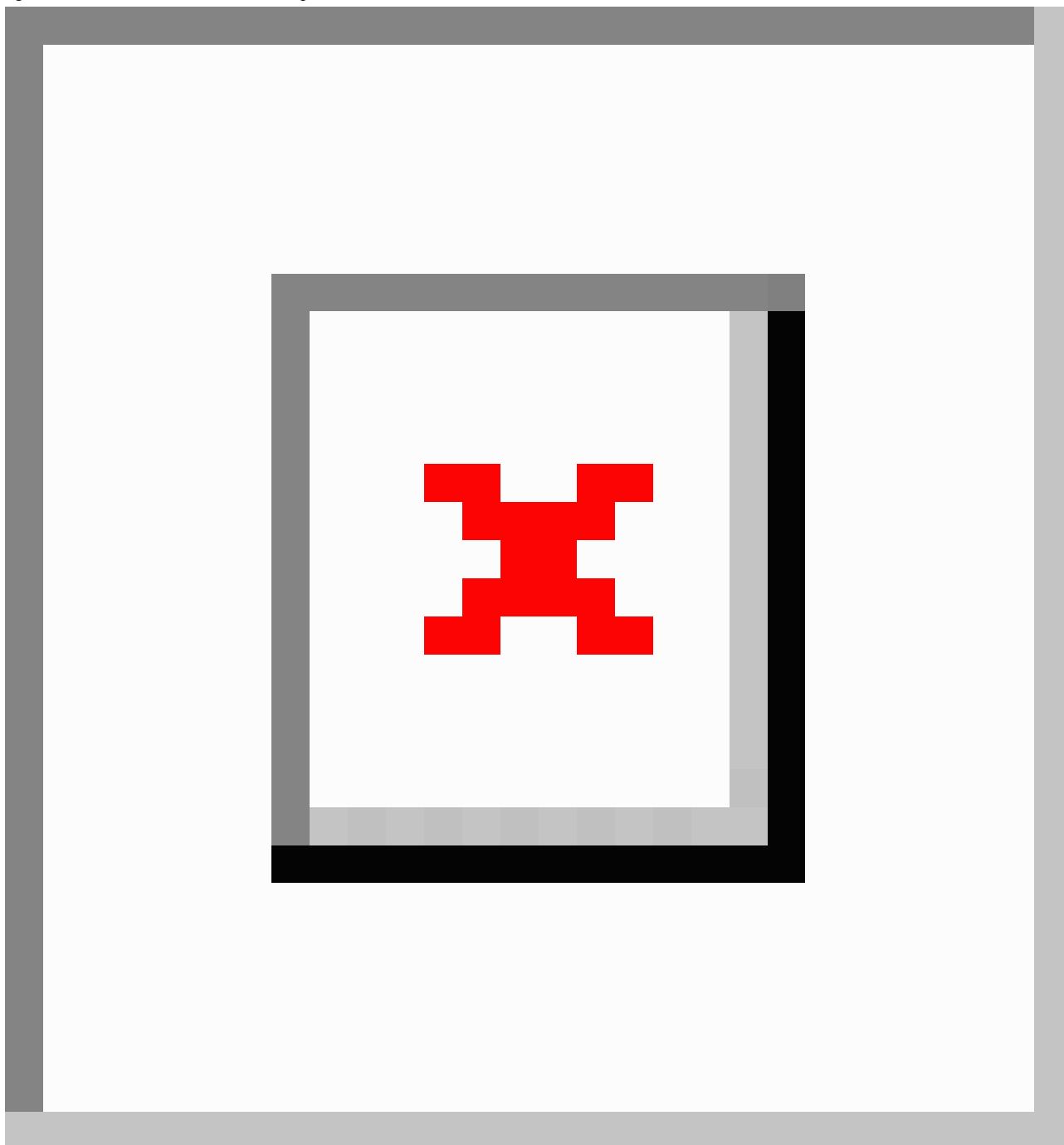
**Table 2.** Assessing Base du Bien Vieillir (BBV) influenza-like illness (ILI) and acute gastroenteritis (AGE) syndromes data transmissions precision during 3 periods: last summer (2016: 33rd week), the Last ILI (2017: 3rd week) and (2017: first week) epidemics weeks' peaks.

| Disease | Epidemic period | Week of study | Residents with syndromes, n | Residents with syndromes x transmission days, n | Residents with influenza-like illness or acute gastroenteritis, n | False positive | True positive | Precision[a] or positive predictive value |
|---------|-----------------|---------------|-----------------------------|--------------------------------------------------|---------------------------------------------------------------------|----------------|---------------|--------------------------------------------|
| ILI | No | sem33-2016 | 5399 | 10,631 | 56 | 3 | 53 | 0.95 |
| AGE | No | sem33-2016 | 5399 | 10,631 | 26 | 2 | 24 | 0.92 |
| ILI | Yes | sem03-2017 | 6013 | 12,215 | 318 | 5 | 313 | 0.98 |
| AGE | Yes | sem01-2017 | 5862 | 11,933 | 68 | 3 | 65 | 0.96 |

[a]Precision=TP/(TP+FP).

### Stability

As detailed in this section, the data transmission stability was evaluated by studying the weekly syndromes' frequencies for every NH from November 1, 2010, as well as the ratio of weeks with syndromic data transmissions:

1. The weekly 26 syndromes frequencies averaged over the number of NH (126) ranged from 21.45 to 180.57 (mean=88.5, SD=22.8).
2. The ratio of weeks of data transmissions per NH was built by computing the number of data transmissions weeks versus the data transmissions weeks span and ranged from 89% to 100% (mean=100%, SD=1%). Only 1 NH had a ratio of less than 95%.

Finally, the syndromic data flow over time studied with 3 chronic illnesses and falls syndromes distributions showed great stability (see Multimedia Appendix 5 for further details).

### Flexibility-Timeliness-Representativeness-Usefulness

Table 2 showed the flexibility and reactiveness of our syndromic system during epidemic periods, following at the same time the increasing number of cases, going from a weekly population of 5399 to 6015 without losing any precision, from 95% to 98%. As shown this winter, by using the NH indexes and regions, the flu epidemic was followed geographically, week after week, starting from the beginning, detecting where the epidemic was most intense, tracking the most severe cases and related hospitalizations and deaths.

Great geographic heterogeneity was detected between regions in terms of ratios of infected; the Rhone valley (Rhone Alpes) and south (Sud) being the most afflicted with 24% and 19% and the southeast (Sud-Est) being the least afflicted with only 9%. There was also great variability in terms of population characteristics, for example, those from the southwest were the oldest afflicted at a mean age of 89.7 years, whereas the southern residents were the youngest at 87.6 years, 2 years and a month difference being quite a lot given the mean duration of residents' stay. Finally, among 1800 residents with a flu transmission, only 1 had no personal data.

### Surveillance Algorithms' Quality

The 3 surveillance package EARS's algorithms as well as Bayes' algorithm were compared for both diseases. For every algorithm, the CI level defines the threshold used to trigger statistical alarms. We used a CI level of 0.001 for EARS_C$i$, $i=1, 2, 3$ and also a CI level of 0.025 for EARS_C3 and Bayes' algorithms. 0.001 is the EARS_C3 default level, whereas 0.025 is the Bayes' algorithm default level (see the algorithm quality in Table 3). The EARS_C3 with alpha=0.025 gave the best results for both diseases (in italics in Table 3). Nevertheless, the Bayes' algorithm seemed better to define alarm weeks when epidemics were less intense as for ILI 2015 to 2016 and AGE 2013 to 2014 seasons, as there were fewer lag weeks between the Bayes' alarm weeks and the Sentinelles outbreak weeks (green arrows were added in Figures 3-6 to highlight this trend).

Either with ILI or AGE TS, mostly coherence between NH data and the Sentinelles data could be witnessed. In addition, only 12 weeks of data (see Table 3) were needed to detect outbreaks, most of the time several weeks ahead of Sentinelles' outbreaks. This was especially true for the last flu season (winter 2016 to 2017 in Figures 3 and 4).

**Table 3.** Comparing the surveillance algorithms' quality on Base du Bien Vieillir (BBV) influenza-like illness (ILI) and acute gastroenteritis (AGE) time series with ILI and AGE Sentinelles' outbreaks detection, by using early aberration reporting system (EARS).

| Disease and algorithm | Number of weeks | True positive | False positive | True negative | False negative | Sensitivity or recall | Specificity | Distance[a] | Precision[b] or positive predictive value |
|---|---|---|---|---|---|---|---|---|---|
| **Influenza-like illness** | | | | | | | | | |
| Bayes[c] | 12 | 22 | 44 | 210 | 32 | 0.407 | 0.827 | 0.617 | 0.333 |
| EARS_C1[d] | 7 | 6 | 12 | 248 | 48 | 0.111 | 0.954 | 0.890 | 0.333 |
| EARS_C2[d] | 9 | 13 | 26 | 232 | 41 | 0.241 | 0.899 | 0.766 | 0.333 |
| EARS_C3[d] | 12 | 19 | 31 | 225 | 35 | 0.352 | 0.879 | 0.659 | 0.380 |
| *EARS_C3* [c] | *12* | *26* | *40* | *216* | *28* | *0.482* | *0.844* | *0.542* | *0.394* |
| **Acute gastroenteritis** | | | | | | | | | |
| Bayes[c] | 12 | 16 | 18 | 251 | 23 | 0.410 | 0.933 | 0.594 | 0.471 |
| EARS_C3[d] | 12 | 16 | 13 | 258 | 23 | 0.410 | 0.952 | 0.607 | 0.592 |
| *EARS_C3* [c] | *12* | *21* | *26* | *245* | *18* | *0.539* | *0.904* | *0.471* | *0.447* |

[a]Distance=sqrt $((1\text{-spec})^2 + (\text{sens-1})^2)$ is the Euclidean distance of (specificity, sensibility) from (1, 1).

[b]Precision=TP/ (TP+FP) is the true positives ratio among the positives.

[c]CI with alpha=.025. Italicization indicates the best results, and therefore, the best method to detect both ILI and AGE.

[d]CI with alpha=.001.

## Discussion

### Principal Findings

We built and assessed a national ecological NH PH SS dedicated to senior citizens. By using a national network of 126 NH and extracting all sociodemographic as well as daily medical data from EHRs, a cohort of 41,061 residents was built. Through textual analysis of clinical narratives (CNs), we implemented ARI_ILI and AGE syndromes. We also engineered related TS by computing weekly head counts. Alarms with EARS_C3 and Bayes algorithms on these, over a 6-year period, allowed us to forecast the 2016 to 2017 influenza outbreak by more than 2 weeks; as can be seen in Figure 5, our statistical alarms were triggered in December, whereas the influenza epidemic according to SPF started only in January.

With just 4 tables, this IS of a new kind showed that it is possible to follow almost every resident every day, where he or she is, during his or her entire NH life, hopefully selecting most of his or her ARI-ILI and AGE health events, from NH entry until death or exit. Furthermore, each relevant syndrome is defined by 2 syndromic representations: either a simple additive syndromic image, that is, its 4 Boolean [65] syndromic components allowing whatever filtering, or its literal expression for further textual analysis or in-depth health questioning. By this whole process, free textual information extracted from CN was shaped into numerical data for further statistical or machine learning analysis.

In this study, we engineered a real NH SSS on qualitative data, offering immediate accessibility without adding any extra work to medical staff [11]. By using SQL-like pattern matching [37] and Delphi-like experts' consensus [57,58] on the data transmissions file, we followed last season ARI-ILI and AGE

epidemics and found almost in real time that the flu dramatically reached NH residents, tracking them geographically and timely, searching for flu-related hospitalizations and deaths. Preventing disruptions of medical tasks and medical and paramedical staff turnover by predicting even 1 or 2 weeks ahead, the epidemic intensity could greatly improve the NH human resources management over time and help to prevent sanitary disasters by strengthening hygiene measures, for example.

As explained in [12], early detection of outbreaks can be achieved in 3 ways: first, by prompt recognition and reporting of disease case reports. Here, we could find most of flu and AGE cases by syndromic descriptions fed in the data transmissions table. Second, by improving the ability to recognize patterns indicative of a possible outbreak early in its course, using analytic tools, counting syndromes by NH, and building time series with the surveillance package. Third, by exploiting data that can signify an outbreak earlier in its course. More specifically, adding hospitalizations and deaths syndromes to the ARI_ILI, AGE syndromes allowed us to assess the flu and AGE outbreaks intensities as well as their severities long before the French health authorities this last season and follow precisely and locally the residents' syndromic population because of the NH and residents' indexes.

This framework with its 3 components, wholly described in Figure 1, has shown its efficacy as a public health SS for early detection of outbreaks. By bringing to light new data not available elsewhere when needed, this SSS improves NH ARI-ILI epidemics' knowledge. Its tools' efficacy could even be quantified by assessing syndromes' precision, stability, flexibility, timeliness, representativeness, and finally algorithms' quality [12,66].

For the AGE data, even with lots of cases, a good correlation could be found for every winter season between the NH alarm weeks and Sentinelles outbreak weeks (as shown in the last row in Table 3 by the small distance value of 0.471). The first ones almost always precede the latter by several weeks, except for the 2014-2015 winter where the AGE epidemic essentially reached senior citizens in NH [85]. During last winter, the AGE outbreak started at the same time as in other NH in France.

## Limitations

This SSS using mostly the transmissions' qualitative data is neither exhaustive as some syndromes may still not be described in the SSS nor complete, as medical staff may not have fed all syndromic information on some day for whatever reason. So, ILI and AGE syndromic data recall, what proportion of cases in classes were correctly assigned to their classes [65], could not be assessed. At this moment, the syndromic information depends essentially on the medical staffs' available time and dedication to feed the system as shown in the Results section Syndromic Data Analysis subsection Stability subsubsection, where 1 NH had a ratio of data transmissions weeks of 89%, with 293 weeks of data transmission over a total span of 329 weeks.

As soon as the cold season begins, elderly people may get a respiratory syncytial virus (RSV), similar to very young children. In fact, RSV is a common cause of acute respiratory illness in older adults as the risk of serious respiratory infection increases with age [86,87]. Usually, RSV spreads quickly just before flu or at the same time and is largely indistinguishable from influenza based on clinical presentation alone [50,51,86]. It is rather a recurring problem in older adults causing 2% to 5% of adult community-acquired pneumonia [88]. Triggering an alarm even for RSV would allow to quickly organize care for the residents.

Then, by following our syndromic ARI_ILI data, 2 trends could be traced, one starting in early November, maybe the RSV, followed by another one later, starting usually in December as this year or later as last year. Depending on the flu epidemic characteristics and as ARI, ILI, and RSV could not be distinguished in our text mining algorithm, a flu threshold could be detected whenever appropriate or several weeks ahead. As can be seen, during the 2013-2014 and 2015-2016 winters, between the first alarm weeks and the outbreak weeks, quite long times elapsed [89,90], but as not really reaching elderly people, there was not something clear to find. However, during the 2014-2015 and 2016-2017 winters, we found a much better correlation between the two, the first ones, probably because of RSV, always preceding the latter by approximately 8 weeks (Figures 3 and 4), thus often triggering alarms before those of the Sentinelles network.

At the same time, we found proportionately much more ILI new cases with our SSS than with the Sentinelles network, especially for this last influenza season (see the last ARI-ILI surge at the beginning of 2017 in Figures 3 and 4). As a type A influenza virus, it reached people older than 75 years much more than the rest of the French population [81]. Then, as soon as clusters of NH ARI cases appeared, many flu tests had to be done to label residents as flu positive or negative. Moreover, even as some

tests were negative, they derived from the flu epidemic health protocol and were mandatory to HRA hygienic safety measures [91-95], increasing the number of cases still more.

Nevertheless, as detailed above, fewer lag weeks were found with the Bayes' algorithm and even an overlap of alarm weeks and outbreak weeks for the ILI 2015-2016 (Figure 3) and AGE 2013-2014 (Figure 5) epidemics and nothing like that with the EARS_C3 algorithm (Figures 4 and 6). We could try in the following years to mix both algorithms as done in the study by Baroukh [96] for Salmonella and decide triggering an alarm whenever 1 of the 2 algorithms reaches its alarm threshold, probably improving both sensitivity and specificity. Alternatively, as in the new MASS (Module for the Analysis of SurSaUD and Sentinelles' data) system [68] designed by SPF, we could combine 3 statistical methods and 3 different data sources, used since January 2016 to define the public health alerts.

Finally, the epidemiologic analysis and interpretation steps (Figure 1) were not fully automated. Some work still needs to be done, especially the whole Sentinelles data extraction process. Some similar job was done before on another project [96,97].

## Conclusions

Outbreak alerts are more reliable when systems focus on specific syndromes that reflect high-probability events such as influenza [62], as could be seen in this real-life experiment. However, there is always room for improvement, as the aggregation of ARI and ILI as well as RSV constraint shows. Nevertheless, this IS gives already a rich and detailed *syndromic* image of these residents. Moreover, as syndromes are modular and the *Pentaho* platform [64] allows extraction from different data silos, it will be possible to add new syndromes, maybe RSV, whenever needed and to adapt them to the new IS that is twice as big and due next year.

This study follows another work on CN using textual analysis and clearing the way for this syndromic health IS design [98]. Tracking flu and AGE epidemics seasons almost in real time and following their impact especially during this last year acute flu season has helped to show the usefulness of this SSS. In addition, the (November 2010-June 2016) syndromic data were used to build ARI_ILI and AGE algorithms, and nothing had to be added or retrieved to follow these last season epidemics' trends, so these algorithms exhibited flexibility, adaptability, stability, and timeliness.

This study highlights some differences between the NH residents' population and the general population, which hampers a better correspondence between NH alarm weeks and Sentinelles outbreak weeks. The main challenges here are extending the syndromic IS, improving the syndromes descriptions, as well as better taking into account NH residents' distinctiveness. Monitoring flu and AGE using the BBV IS could give way to a real SS for all senior citizens in France. For example, there are incoming discussions between Korian and HRA about targeting RSV besides flu and handling what differentiates them.

Korian NHs are already working with HRA at a local level, exchanging clinical data with them whenever outbreaks are

detected. This data sharing could then be extended with syndromic data integration, resulting in HRA reactiveness improvement [99]. Indeed, syndromic data are always available before, even if less precise. NH residents as a whole are a frail and captive population functioning as an ever-increasing reservoir for any contagious illness [100,101]. It is then essential to be able to prevent with all possible disposable tools any health catastrophe in the near future.

This syndromic IS offers a real opportunity for finding new ways to seniors' functioning modelization and opens, hopefully, the path toward specific clinical hypotheses formulation. Other works included studying the use of this IS applied to other public health problems such as frequent falls or falls with casualties [102] but also working toward a better life ending with cancer [103]. Ultimately, the aims are removing all preventable deaths and improving the residents' end of life with more autonomy, less pain, and an improved quality of life, translating this new knowledge into health benefits for seniors everywhere.

## Authors' Contributions

TD and LJ had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors contributed to the study concept and design, critically revised the manuscript for important intellectual content, and supervised the conduct of the study. TD oversaw the data extraction and analysis as well as statistical analysis, and LJ obtained funding.

## Conflicts of Interest

LJ does not have any financial competing interests to report, but he is a member of the Korian Aging Well Committee.

## Multimedia Appendix 1

The Korian nursing home network in France.

[PDF File (Adobe PDF File), 81KB-Multimedia Appendix 1]

## Multimedia Appendix 2

The anonymization process.

[PDF File (Adobe PDF File), 19KB-Multimedia Appendix 2]

## Multimedia Appendix 3

The Base du Bien Vieillir (BBV) acute respiratory infection and influenza-like illness and acute gastroenteritis syndromic information building process in 4 phases: the 4 BBV tables, 2 syndromic examples, and the BBV 26 syndromes list.

[PDF File (Adobe PDF File), 93KB-Multimedia Appendix 3]

## Multimedia Appendix 4

How public health alerts are defined: statistical alerts according to the Sentinelles network; true alerts according to the MASS system; and how the surveillance package works.

[PDF File (Adobe PDF File), 57KB-Multimedia Appendix 4]

## Multimedia Appendix 5

Details about syndromic data-flow stability rating syndromic data flow stability over time with 4 syndromes frequencies: diabetes, cardiovascular problems, depression, and frequent falls.

[PDF File (Adobe PDF File), 52KB-Multimedia Appendix 5]

## References

1.  Luxembourg: Publications Office of the European Union. European Commission. 2017 Oct. The 2018 Ageing Report Underlying Assumptions and Projection Methodologies URL: https://ec.europa.eu/info/sites/info/files/economy-finance/ip065_en.pdf [accessed 2017-12-28] [WebCite Cache ID 6w377yoQH]

2.  Desrivierre D. Institut national de la statistique et des études économiques. 2017 Jun. [By 2050, the population would increase in all metropolitan areas] URL: https://www.insee.fr/fr/statistiques/2867738 [accessed 2017-12-28] [WebCite Cache ID 6w38s0ipq]

3.  Delbès C, Gaymu J. Revue Gerontologie et Societe. 2005. [The population in EHPAD in France Who lives in an institution?] URL: https://www.cairn.info/revue-gerontologie-et-societe1-2005-1-page-13.htm [accessed 2018-01-03]

4.  Monaghan P, Charmantier A, Nussey DH, Ricklefs RE. The evolutionary ecology of senescence. Funct Ecol 2008 Jun;22(3):371-378 [FREE Full text] [doi: 10.1111/j.1365-2435.2008.01418.x]

5.  Berge GT. Association for Information Systems. 2016. Drivers and barriers to structuring information in Electronic Health Records (2016) URL: http://aisel.aisnet.org/pacis2016/18/ [accessed 2017-12-28] [WebCite Cache ID 6w3AQiskr]

6.  Institut national de la santé et de la recherche médicale [National Institute for Health and Medical Research]. 2011 Nov. [Dependency project 4 epidemiological cohorts Upper Normandy, Paquid, 3 Cities and AMI] URL: http://www.cnsa.fr/documentation/_projet_dependance_4_cohortes_cnsa_version_finale_nov2011_.pdf [accessed 2018-01-13] [WebCite Cache ID 6wRmUQsvL]

7.  Banks J, Batty GD, Nazroo J, Steptoe A. English Longitudinal Study of Ageing. 2016 Oct. The dynamics of ageing.: Evidence from the English Longitudinal Study of Ageing 2002-15 (Wave 7) URL: http://www.elsa-project.ac.uk/publicationDetails/id/8696 [accessed 2018-01-13] [WebCite Cache ID 6wRmtF4UA]

8.  The Survey of Health, Ageing and Retirement in Europe. SHARE - Survey of Health, Ageing and Retirement in Europe URL: http://www.share-project.org/home0.html [accessed 2018-01-13]

9.  Urban Institute. Final Report: Lessons from the Literature on Electronic Health Record Implementation URL: https://www.healthit.gov/sites/default/files/hit_lessons_learned_lit_review_final_08-01-2013.pdf [accessed 2017-12-28] [WebCite Cache ID 6w3D5k2iW]

10. Institut national de la statistique et des études économiques. 2017. [Tables of the French economy] URL: https://www.insee.fr/fr/statistiques/2569388?sommaire=2587886 [accessed 2018-01-13] [WebCite Cache ID 6wRoVTec3]

11. Centers for Disease Control and Prevention. 2001. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm [accessed 2017-12-28] [WebCite Cache ID 6w3EbepiT]

12. Centers for Disease Control and Prevention. 2004 Sep 24. Overview of Syndromic Surveillance: What is Syndromic Surveillance? URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/su5301a3.htm [accessed 2017-12-28] [WebCite Cache ID 6w3F1H5NA]

13. Centers for Disease Control and Prevention. Framework for Evaluating Public Health Surveillance Systems for Early Detection of Outbreaks URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5305a1.htm [accessed 2017-12-28] [WebCite Cache ID 6w3FR4fq3]

14. Katz R, May L, Baker J, Test E. Redefining syndromic surveillance. J Epidemiol Glob Health 2011 Dec;1(1):21-31 [FREE Full text] [doi: 10.1016/j.jegh.2011.06.003] [Medline: 23856373]

15. Josseran L, Fouillet A. [Syndromic surveillance: review and prospect of a promising concept]. Rev Epidemiol Sante Publique 2013 Apr;61(2):163-170. [doi: 10.1016/j.respe.2013.01.094] [Medline: 23481885]

16. Fouillet A, Medina S, Medeiros H. Bulletin Épidémiologique Hebdomadaire (BEH). 2014 Jan 21. [Syndromic Surveillance in Europe: the Triple-S European Project] URL: http://invs.santepubliquefrance.fr//beh/2014/3-4/index.html [accessed 2017-12-28] [WebCite Cache ID 6w3LByIfI]

17. Soulakis ND. University of Pittsburgh. 2012. Syndromic Surveillance for Bioterrorism Related Inhalation Anthrax in an Emergency Department Population URL: http://d-scholarship.pitt.edu/17139/ [accessed 2017-12-29] [WebCite Cache ID 6w54b0a6h]

18. Flamand C, Larrieu S, Couvy E, Jouves B, Josseran L, Filleul L. Validation of a syndromic surveillance system using a general practitioner house calls network, Bordeaux, France. Euro Surveill 2008 Jun 19;13(25) [FREE Full text] [Medline: 18761939]

19. Josseran L, Caillère N, Brun-Ney D, Rottner J, Filleul L, Brucker G, et al. Syndromic surveillance and heat wave morbidity: a pilot study based on emergency departments in France. BMC Med Inform Decis Mak 2009 Feb 20;9:14 [FREE Full text] [doi: 10.1186/1472-6947-9-14] [Medline: 19232122]

20. Van Ganse E E, Belhassen M. REG 2015 Winter Summit. 2015 Jan 23. SNIIRAM: Primary and Secondary Case Resource Use in France URL: https://fr.slideshare.net/RespiratoryEffectivenessGroup/sniiram-primary-and-secondary-care-resource-use-in-france [accessed 2018-10-28] [WebCite Cache ID 73VtqOj9D]

21. Moulis G, Lapeyre-Mestre M, Palmaro A, Pugnet G, Montastruc JL, Sailler L. [French health insurance databases: what interest for medical research?]. Rev Med Interne 2015 Jun;36(6):411-417 [FREE Full text] [doi: 10.1016/j.revmed.2014.11.009] [Medline: 25547954]

22. Sentinelles Network. [Metropolitan France] URL: https://websenti.u707.jussieu.fr/sentiweb/ [accessed 2018-10-28] [WebCite Cache ID 73VycCjIE]

23. Caillère N, Fouillet A, Henry V. Sante publique France. 2012. The French Health Surveillance System for Emergencies and Deaths (SurSaUD®) URL: http://invs.santepubliquefrance.fr/Publications-et-outils/Rapports-et-syntheses/Autres-thematiques/2012/Le-systeme-francais-de-Surveillance-sanitaire-des-urgences-et-des-deces-SurSaUD-R

24. Mathieu A, Larras B, Leroy JP, Chamberland G, Benhalima B, Ruello M, et al. Sante publique France. 2015 Feb. [What syndromic surveillance system is based on health surveillance in Normandy?] URL: http://invs.santepubliquefrance.fr/content/download/132209/472872/version/33/file/bvs_normandie_24_decembre_2016.pdf

25. Smith S, Smith GE, Olowokure B, Ibbotson S, Foord D, Maguire H, et al. Early spread of the 2009 influenza A(H1N1) pandemic in the United Kingdom--use of local syndromic data, May-August 2009. Euro Surveill 2011 Jan 20;16(3) [FREE Full text] [Medline: 21262185]

26. Josseran L, Fouillet A, Caillère N, Brun-Ney D, Ilef D, Brucker G, et al. Assessment of a syndromic surveillance system based on morbidity data: results from the Oscour network during a heat wave. PLoS One 2010 Aug 09;5(8):e11984 [FREE Full text] [doi: 10.1371/journal.pone.0011984] [Medline: 20711252]

27. Ducoudray JM, Eon Y, Le Provost C. Caisse nationale de solidarité pour l'autonomie. 2017. [The PATHOS model, User Guide 2017] URL: http://www.cnsa.fr/documentation/modele_pathos_2017.pdf [accessed 2017-12-29] [WebCite Cache ID 6w4ZRwZMi]

28. Portail national d'information pour l'autonomie des personnes âgées et l'accompagnement de leurs proches. 2018 May 22. [How is the GIR determined?] URL: http://www.pour-les-personnes-agees.gouv.fr/beneficier-daides/lallocation-personnalisee-dautonomie-apa/comment-le-gir-est-il-determine [accessed 2017-12-29] [WebCite Cache ID 6w4a5soTN]

29. Closon MC, Habimana L, Laokri S. Ministère des Affaires sociales et de la Santé, Paris. 2006. [The AGGIR PATHOS SOCIOS Model: a Potential Instrument for Funding, Programming and Internal Management of Geriatric and Rehabilitation Services] URL: http://www.medcomip.fr/region/region-gir-pathos/rapport-novella-cs-2012.pdf [accessed 2017-12-29] [WebCite Cache ID 6w4b80vvj]

30. Neirynck I, Closon MC, Swine C. Ministère des Affaires sociales et de la Santé, Paris. 2006. [Validation tests of the AGGIR PATHOS SOCIOS model in geriatric and rehabilitative services] URL: http://www.medcomip.fr/region/region-gir-pathos/rapport-novella-cs-2012.pdf [accessed 2017-12-29] [WebCite Cache ID 6w4b80vvj]

31. Dain L. Psychomotricity - Faculties of Medicine of Toulouse. 2011. [What can psychomotricity bring to dependent elderly people? Illustration with 2 case studies] URL: http://www.psychomot.ups-tlse.fr/Dain2011.pdf [accessed 2017-12-29] [WebCite Cache ID 6w4c1bd5J]

32. Hazif-Thomas C, Reber C, Bonvalot T. Dysexecutive syndrome and late depression. Ann Méd Psychol 2005 Sep;163(7):569-576. [doi: 10.1016/j.amp.2005.07.005]

33. Cohen R, Elhadad M, Elhadad N. Redundancy in electronic health record corpora: analysis, impact on text mining performance and mitigation strategies. BMC Bioinformatics 2013 Jan 16;14:10 [FREE Full text] [doi: 10.1186/1471-2105-14-10] [Medline: 23323800]

34. Bui AA, Taira RK, El-Saden S, Dordoni A, Aberle DR. Automated medical problem list generation: towards a patient timeline. Stud Health Technol Inform 2004;107(Pt 1):587-591. [doi: 10.3233/978-1-60750-949-3-587] [Medline: 15360880]

35. Burton MM, Simonaitis L, Schadow G. Medication and Indication Linkage: A Practical Therapy for the Problem List? In: AMIA 2008 Symposium Proceedings. 2008 Presented at: AMIA 2008 Symposium; 2008; Savannah, Georgia p. 86-90 URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655999/

36. Campbell JR. Strategies for problem list implementation in a complex clinical enterprise. Proc AMIA Symp 1998:285-289 [FREE Full text] [Medline: 9929227]

37. tutorialspoint. SQL-LIKE clause URL: https://www.tutorialspoint.com/sql/sql-like-clause.htm [accessed 2018-10-29] [WebCite Cache ID 6w4gGA4em]

38. Adam S, Bonsang E, Grotz C, Perelman S. Occupational activity and cognitive reserve: implications in terms of prevention of cognitive aging and Alzheimer's disease. Clin Interv Aging 2013;8:377-390 [FREE Full text] [doi: 10.2147/CIA.S39921] [Medline: 23671387]

39. Lebert F, Leroy M, Pasquier F, Strubel D. Young onset demented patients in French cognitive-behavioral specialized units. Geriatr Psychol Neuropsychiatr Vieil 2016 Jun 01;14(2):194-200. [doi: 10.1684/pnv.2016.0607] [Medline: 27277152]

40. Abadie R, Voisin T. Groupe Fmc Gériatrie et médecins coordonnateurs 31. 2014. [Procedure for the detection and management of behavioral psycho symptoms of dementia in nursing homes] URL: http://www.medcomip.fr/region/region-outils/outils-egs/troubles-cpmt/procedure-spcd-ehpad.pdf [accessed 2017-12-29] [WebCite Cache ID 6w4hHAHmN]

41. World Health Organization. 2016 Apr. WHO Mental health and older adults Fact sheets URL: http://www.who.int/mediacentre/factsheets/fs381/en/ [accessed 2017-12-29] [WebCite Cache ID 6w4lVpnZM]

42. de Villiers L. Continuing Medical Education. Frailty URL: http://www.cmej.org.za/index.php/cmej/article/view/2868/3235 [accessed 2017-12-29] [WebCite Cache ID 6w4mPVDFL]

43. World Health Organization. 2016 Sep. [WHO Falls Quick Reference] URL: http://www.who.int/mediacentre/factsheets/fs344/fr/%20 [accessed 2017-12-29] [WebCite Cache ID 6w4mjnzJ2]

44. Rubenstein LZ, Josephson KR, Robbins AS. Falls in the nursing home. Ann Intern Med 1994 Sep 15;121(6):442-451. [doi: 10.7326/0003-4819-121-6-199409150-00009] [Medline: 8053619]

XSL•FO
RenderX

45. Arai H, Ouchi Y, Yokode M, Ito H, Uematsu H, Eto F, Members of Subcommittee for Aging. Toward the realization of a better aged society: messages from gerontology and geriatrics. Geriatr Gerontol Int 2012 Jan;12(1):16-22. [doi: 10.1111/j.1447-0594.2011.00776.x] [Medline: 22188494]

46. Zaslavsky O, Thompson H, Demiris G. The role of emerging information technologies in frailty assessment. Res Gerontol Nurs 2012 Jul;5(3):216-228. [doi: 10.3928/19404921-20120410-02] [Medline: 22533942]

47. Société Française de Gériatrie et de Gérontologie. International Association of Gerontology and Geriatrics - Société Française de Gériatrie et de Gérontologie. 2015 Mar 02. [Identifying and maintaining the independence of frail elderly people] URL: https://fragilite.org/livre-blanc.php [accessed 2017-12-29]

48. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [Medline: 11253156]

49. Meystre S, Haug PJ. Automation of a problem list using natural language processing. BMC Med Inform Decis Mak 2005 Aug 31;5:30 [FREE Full text] [doi: 10.1186/1472-6947-5-30] [Medline: 16135244]

50. Lindsay K. Physicians Weekly. 2016 Sep 02. Managing Adult Respiratory Syncytial Virus URL: https://www.physiciansweekly.com/managing-adult-respiratory-syncytial-virus/ [accessed 2018-10-29] [WebCite Cache ID 73X5B5T7j]

51. National Center for Immunization and Respiratory Diseases (NCIRD), Division of Viral Diseases. Centers for Disease Control and Prevention. 2018 Jun 26. Respiratory Syncytial Virus Infection (RSV) URL: https://www.cdc.gov/rsv/index.html [accessed 2018-10-29] [WebCite Cache ID 73X5nEvk1]

52. Wagner MM, Moore AA, Aryel RM. Natural Language Processing for Biosurveillance. In: Handbook of Biosurveillance. Burlington: Elsevier Science; 2006.

53. Fieschi M, Bouhaddou O, Beuscat R. Computer Science at the Service of the Patient: Financial Statements. Paris: Springer-Verlag; May 2000.

54. Kutais BG. Full-Text Search in Electronic Health Recrods: Challenges And Oppurtunities. In: Internet Policies & Issues: V. 7. New York, United States: Nova Science Pub Inc; 2018.

55. Dick B. International Congress of Action Research and Process Management, Griffith University, Brisbane. 1990 Jul. Dialectical processes URL: http://www.aral.com.au/DLitt/DLitt_P24delphi.pdf [accessed 2018-10-29] [WebCite Cache ID 73X6Xz1AW]

56. Debin M, Souty C, Turbelin C, Blanchon T, Boëlle PY, Hanslik T, DelFluWeb Study Group. Determination of French influenza outbreaks periods between 1985 and 2011 through a web-based Delphi method. BMC Med Inform Decis Mak 2013 Dec 24;13:138 [FREE Full text] [doi: 10.1186/1472-6947-13-138] [Medline: 24364926]

57. Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. J Clin Epidemiol 2003 Dec;56(12):1150-1156. [Medline: 14680664]

58. Chapman WW, Dowling JN, Baer A, Buckeridge DL, Cochrane D, Conway MA, et al. Developing syndrome definitions based on consensus and current use. J Am Med Inform Assoc 2010 Sep;17(5):595-601 [FREE Full text] [doi: 10.1136/jamia.2010.003210] [Medline: 20819870]

59. Liu F, Chen J, Jagannatha A. Arxiv. Learning for Biomedical Information Extraction Methodological Review of Recent Advances URL: https://arxiv.org/abs/1606.07993 [accessed 2018-01-01] [WebCite Cache ID 6w9Bmjq3t]

60. Tellier I. Laboratoire Lattice - CNRS. [Introduction to NLP and Linguistic Engineering] URL: http://www.lattice.cnrs.fr/sites/itellier/poly_info_ling/info-ling.pdf [accessed 2018-01-01] [WebCite Cache ID 6w9CUH0tQ]

61. Zhai C, Massung S. Morgan & Claypool Publishers. 2016. Text Data Management and Analysis URL: http://www.morganclaypoolpublishers.com/catalog_Orig/samples/9781970001174_sample.pdf [accessed 2018-01-01] [WebCite Cache ID 6w9Csma5Q]

62. Chretien J, Tomich NE, Gaydos JC, Kelley PW. Real-time public health surveillance for emergency preparedness. Am J Public Health 2009 Aug;99(8):1360-1363. [doi: 10.2105/AJPH.2008.133926] [Medline: 19542047]

63. Andersson MG, Faverjon C, Vial F, Legrand L, Leblond A. Using Bayes' rule to define the value of evidence from syndromic surveillance. PLoS One 2014;9(11):e111335 [FREE Full text] [doi: 10.1371/journal.pone.0111335] [Medline: 25364823]

64. Hitachi Vantara Community. Data Integration - Kettle URL: http://community.pentaho.com/projects/data-integration/ [accessed 2018-10-29] [WebCite Cache ID 73XDJzIzp]

65. Gibbons C, Richards S, Valderas JM, Campbell J. Supervised ML algorithm can classify open-text feedback of doctor performance with human-level accuracy. J Med Internet Res 2017 Dec 15;19(3):e65 [FREE Full text] [doi: 10.2196/jmir.6533] [Medline: 28298265]

66. Höhle M. An R package for the monitoring of infectious diseases. Comput Stat 2007 Aug 15;22(4):571-582. [doi: 10.1007/s00180-007-0074-8]

67. Höhle M, Meyer S, Held L. Cran - R Project. 2018 Jul 25. Package 'surveillance' URL: https://cran.r-project.org/web/packages/surveillance/surveillance.pdf [accessed 2018-10-29] [WebCite Cache ID 73XE2VMXM]

68. Pelat C, Bonmarin I, Ruello M, Fouillet A, Caserio-Schönemann C, Levy-Bruhl D, Regional Influenza study group. Improving regional influenza surveillance through a combination of automated outbreak detection methods: the 2015/16 season in France. Euro Surveill 2017 Dec 10;22(32) [FREE Full text] [doi: 10.2807/1560-7917.ES.2017.22.32.30593] [Medline: 28816649]

69. Institut de veille sanitaire. 2011. La veille et l'alerte sanitaires en France URL: http://solidarites-sante.gouv.fr/IMG/pdf/Rapport_veille_alerte_sanitaire_France.pdf [accessed 2018-05-14] [WebCite Cache ID 6zPFnbyLa]

70. Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. Public Health Rep 1963 Jun;78(6):494-506 [FREE Full text] [Medline: 19316455]

71. Costagliola D, Flahault A, Galinec D, Garnerin P, Menares J, Valleron AJ. A routine tool for detection and assessment of epidemics of influenza-like syndromes in France. Am J Public Health 1991 Jan;81(1):97-99. [Medline: 1983924]

72. Höhle M, Riebler A, Paul M. Cran - R Project. 2007 Nov 17. Getting Started With OutBreak Detection URL: https://cran.r-project.org/web/packages/surveillance/vignettes/surveillance.pdf [accessed 2018-01-01] [WebCite Cache ID 6w9Em7JTU]

73. Cowling BJ, Wong IO, Ho L, Riley S, Leung GM. Methods for monitoring influenza surveillance data. Int J Epidemiol 2006 Oct;35(5):1314-1321. [doi: 10.1093/ije/dyl162] [Medline: 16926216]

74. Centers for Disease Control and Prevention. 2018. CDC Surveillance Resource Center URL: https://www.cdc.gov/surveillancepractice/ [accessed 2018-10-29] [WebCite Cache ID 73XHZF0cB]

75. Yang P, Duan W, Lv M, Shi W, Peng X, Wang X, et al. Review of an influenza surveillance system, Beijing, People's Republic of China. Emerg Infect Dis 2009 Oct;15(10):1603-1608 [FREE Full text] [doi: 10.3201/eid1510.081040] [Medline: 19861053]

76. Jung N. Dépôt des mémoires universitaires après soutenance (DUMAS). 2010. Surveillance sanitaire à partir de données des services d'urgence: modélisation de séries temporelles et analyse automatique URL: https://dumas.ccsd.cnrs.fr/dumas-00516268 [accessed 2018-10-29] [WebCite Cache ID 73XK301ZU]

77. Tokars JI, Burkom H, Xing J, English R, Bloom S, Cox K, et al. Enhancing time-series detection algorithms for automated biosurveillance. Emerg Infect Dis 2009 Apr;15(4):533-539 [FREE Full text] [doi: 10.3201/eid1504.080616] [Medline: 19331728]

78. Cran - R Project. 2018 Oct 25. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics URL: https://cran.r-project.org/web/packages/ggplot2/index.html [accessed 2018-10-29] [WebCite Cache ID 73XKVGRqZ]

79. Cleveland WS, Grosse E, Shyu W. Github.: Chambers JM and Hastie TJ, Wadsworth & Brooks/Cole1992; 1992. Local regression models URL: http://rafalab.github.io/pages/754/section-03.pdf [accessed 2018-01-01] [WebCite Cache ID 6w9E07zjA]

80. Cameron W, Neu A, Murray E, Soetebier K, Cookson S. Advances in Disease Surveillance. 2007. Responding to Syndromic Surveillance Alerts: An Adaptable Protocol for Georgia Health Districts URL: https://www.researchgate.net/publication/266093117_Responding_to_Syndromic_Surveillance_Alerts_An_Adaptable_Protocol_for_Georgia_Health_Districts [accessed 2018-04-04] [WebCite Cache ID 6yQttWi9M]

81. Sante publique France. Bulletin épidémiologique grippe, semaine 9. Saison 2016-2017 URL: http://invs.santepubliquefrance.fr/Dossiers-thematiques/Maladies-infectieuses/Maladies-a-prevention-vaccinale/Grippe/Grippe-generalites/Donnees-de-surveillance/Archives/Bulletin-epidemiologique-grippe-semaine-9.-Saison-2016-2017 [accessed 2018-01-01] [WebCite Cache ID 6w9W2JEgY]

82. Lee DC, Long J, Wall S, Carr BG, Satchell SN, Braithwaite RS, et al. Determining chronic disease prevalence in local populations using emergency department surveillance. Am J Public Health 2015 Sep;105(9):e67-e74. [doi: 10.2105/AJPH.2015.302679] [Medline: 26180983]

83. Gault G, Larrieu S, Durand C, Josseran L, Jouves B, Filleul L. Performance of a syndromic system for influenza based on the activity of general practitioners, France. J Public Health (Oxf) 2009 Jun;31(2):286-292. [doi: 10.1093/pubmed/fdp020] [Medline: 19269992]

84. Meystre S, Haug PJ. Natural language processing to extract medical problems from electronic clinical documents: performance evaluation. J Biomed Inform 2006 Dec;39(6):589-599 [FREE Full text] [doi: 10.1016/j.jbi.2005.11.004] [Medline: 16359928]

85. Septfons A, Barataud D, Chiron E. Sante publique France. 2016 Jun 21. [Surveillance of AGEs in communities for the elderly, National assessment of five winter monitoring seasons (November 2010-May 2015)] URL: http://invs.santepubliquefrance.fr//beh/2016/18-19/2016_18-19_2.html [accessed 2018-01-01] [WebCite Cache ID 6w9SI7s0v]

86. National Foundation for Infectious Diseases. 2016 Sep. Respiratory Syncytial Virus in Older Adults: A Hidden Annual Epidemic A Report URL: http://www.nfid.org/publications/reports/rsv-report.pdf [accessed 2018-10-29] [WebCite Cache ID 73XMzxSr0]

87. Falsey AR, Hennessey P, Formica M, Cox C, Walsh EE. Respiratory syncytial virus infection in elderly and high-risk adults. N Engl J Med 2005 Apr 28;352(17):1749-1759. [doi: 10.1056/NEJMoa043951] [Medline: 15858184]

88. Falsey AR, Walsh EE. Respiratory syncytial virus infection in adults. Clin Microbiol Rev 2000 Jul;13(3):371-384 [FREE Full text] [doi: 10.1128/CMR.13.3.371] [Medline: 10885982]

89. Sante publique France. 2014. [Influenza Surveillance in Metropolitan France 2013-2014 Season] URL: http://invs.santepubliquefrance.fr/Publications-et-outils/BEH-Bulletin-epidemiologique-hebdomadaire/Archives/2014/BEH-n-28-2014 [accessed 2018-01-01] [WebCite Cache ID 6w9SXCYlh]

90. Sante publique France. 2016. [Influenza Surveillance in Metropolitan France 2015-2016 Season] URL: http://invs.santepubliquefrance.fr/beh/2016/32-33/pdf/2016_32-33.pdf [accessed 2018-01-01] [WebCite Cache ID 6w9T4F6WY]

91.  Ministry of Solidarity and Health. 2011 Oct. [National Influenza Pandemic Prevention and Control Plan 2011 Help document for preparation and decision] URL: http://solidarites-sante.gouv.fr/IMG/pdf/Plan_Pandemie_Grippale_2011.pdf [accessed 2018-01-01] [WebCite Cache ID 6w9WFMtRW]

92.  Institut de veille sanitaire. 2014 Sep. [Report Card for Group Cases of Acute Respiratory Infections (ARI) in the InVS ARS Seniors Community] URL: http://www.iledefrance.paps.sante.fr/fileadmin/ILE-DE-FRANCE/PAPS/Informations_pratiques/MDO/PA_en_IRA/FICHE_DE_SIGNALEMENT__IRA.pdf [accessed 2018-01-01] [WebCite Cache ID 6w9WmJOay]

93.  Syndicat National des Établissements et Résidences privés pour Personnes Âgées. 2016 Oct 06. [Practical guide for prevention and control in case of influenza pandemic in EHPAD] URL: http://www.cpias-ile-de-france.fr/docprocom/doc/Synerpa_Guide_grippe_2016.pdf [accessed 2018-01-01] [WebCite Cache ID 6w9XLF0Lk]

94.  Institut Pasteur. [Influenza Fact Sheets] URL: https://www.pasteur.fr/fr/centre-medical/fiches-maladies/grippe [accessed 2018-10-29] [WebCite Cache ID 6w9XmBiHP]

95.  France 3 Center-Val de Loire. [Influenza epidemic: the ARS renews its prevention recommendations] URL: https://france3-regions.francetvinfo.fr/centre-val-de-loire/epidemie-grippe-ars-renouvelle-ses-recommandations-prevention-1170701.html [accessed 2018-01-01] [WebCite Cache ID 6w9Y4tLVr]

96.  Baroukh T. Agence française de sécurité sanitaire des aliments. 2008 Sep. [Development of an unusual event detection system for the surveillance of non-human Salmonella by statistical methods of time series analysis] URL: https://tinyurl.com/yb5b2xxu [accessed 2018-01-02]

97.  Danan C, Baroukh T, Moury F, Jourdan-DA Silva N, Brisabois A, LE Strat Y. Automated early warning system for the surveillance of Salmonella isolated in the agro-food chain in France. Epidemiol Infect 2011 May;139(5):736-741. [doi: 10.1017/S0950268810001469] [Medline: 20598207]

98.  Delespierre T, Denormandie P, Bar-Hen A, Josseran L. Empirical advances with text mining of electronic health records. BMC Med Inform Decis Mak 2017 Aug 22;17(1):127 [FREE Full text] [doi: 10.1186/s12911-017-0519-0] [Medline: 28830417]

99.  Simon N. HIT Consultant. Why Does Clinical Health Data Exchange Remain Such A Struggle? URL: https://hitconsultant.net/2017/03/06/clinical-health-data-exchange-byline/ [accessed 2018-04-13] [WebCite Cache ID 6ye6oD6KT]

100. Kissling E, Rondy M, I-MOVE/I-MOVE+ study team. Early 2016/17 vaccine effectiveness estimates against influenza A(H3N2): I-MOVE multicentre case control studies at primary care and hospital levels in Europe. Euro Surveill 2017 Dec 16;22(7) [FREE Full text] [doi: 10.2807/1560-7917.ES.2017.22.7.30464] [Medline: 28230524]

101. Dorrington MG, Bowdish DM. Immunosenescence and novel vaccination strategies for the elderly. Front Immunol 2013;4:171 [FREE Full text] [doi: 10.3389/fimmu.2013.00171] [Medline: 23825474]

102. Delespierre T, Denormandie P, Josseran L. The Journal of Nursing Home Research. 2016 Nov. URL: http://www.jnursinghomeresearch.com/all-issues.html?a=2016&n=01 [accessed 2018-01-02] [WebCite Cache ID 6wAYRitWM]

103. Delespierre T, Denormandie P, Armaingaud D. Revue d Épidémiologie et de Santé Publique. 2016. [New data and new methods to evaluate the primary care of cancer patients in EHPAD] URL: https://tinyurl.com/yab6r3ao [accessed 2018-01-02] [WebCite Cache ID 6wAZURk3l]

## Abbreviations

**AGE:**  acute gastroenteritis

**ARI:**  acute respiratory infection

**ARI-ILI:**  acute respiratory infection and influenza-like illness

**BBV:**  Base du Bien Vieillir

**CDC:**  Centers for Disease Control and Prevention

**CN:**  clinical narrative

**CUSUM:**  cumulative sums

**EARS:**  early aberration reporting system

**EHR:**  electronic health record

**ETL:**  extract, transform, and load

**FN:**  false negative

**FP:**  false positive

**GIR:**  groupe ISO ressources - ISO resources group

**GP:**  general practitioner

**HRA:**  Health Regional Agencies

**ILI:**  influenza-like illness

**IS:**  information system

**MSL-TM:**  multistep learning and text mining

**NH:**  nursing homes

**PERMF:**  personal electronic resident medical file

**PH:**  public health

**PPV:** positive predictive value
**RSV:** respiratory syncytial virus
**SPF:** Santé publique France
**SQL:** standard query language
**SS:** surveillance system
**SSS:** syndromic surveillance system
**SurSaUD:** Sanitary Surveillance of Urgencies and Deaths
**TN:** true negative
**TP:** true positive
**TS:** time series

XSL•FO

**RenderX**